Applying Bailer's Method for AUC Confidence Intervals to Sparse Sampling

Jerry R. Nedelman,1,3 Ekaterina Gibiansky,1 and David T. W. Lau2

Received March 11, 1994; accepted July 26, 1994

Bailer (1) developed a method for constructing confidence intervals for areas under the concentration-vs-time curve (AUC's) with only one sample per subject but with multiple subjects sampled at each of several time points post dose. We have modified this method to account for estimation of the variances. How the need to estimate variances affects study design is discussed. An extension of Bailer's method is proposed where variances are modeled as a function of the means, in order to get more precise estimates of variances. The modified and extended methods are applied to a rat toxicokinetic study with only two rats per time point per treatment group.

KEY WORDS: area under the curve; degrees of freedom; Satterthwaite's approximation; variance function; toxicokinetics.

INTRODUCTION

Bailer (1) described a technique for estimating the mean area under the curve (AUC) of drug concentration vs time (Cxt) when only one sample per subject is available but with multiple subjects sampled at each of several time points post dose. He also demonstrated how to estimate the standard error of the estimated AUC and how to test statistically for the equality of two mean AUC's or construct a confidence interval for the difference of two mean AUC's.

Bailer's method is elegant and simple. The tests and confidence intervals do not depend on any models for pharmacokinetic response, nor on any assumptions about variance homogeneity. They depend on the assumption of normally distributed data, and as applied by Bailer, they assume that the variances are known.

Whereas Bailer considered AUC's computed between two finite time points, Yuan (2) showed how to extend his method to infinite time, provided one has an estimate of terminal elimination rate.

We are interested in applying Bailer's method, without Yuan's extension, to rodent toxicity studies in drug development (3). In such studies, the toxicokinetic objective is to quantify drug exposure and relate exposure to dose, sex, and duration of dosing. Typically, a treatment group consists of ten or fewer animals, with only one blood sample collected per animal per day. By sampling blood from different animals at different times post dose, the CxT curve can be characterized. Bailer's method can be applied if two or more animals within each treatment group are sampled at each time point. With only ten animals per group, however, replicating at each time point severely restricts the number of time points. We typically use a design with two animals at each of five time points.

This report is about the utility of Bailer's method for such a sparse design. Our investigation revealed a complication of Bailer's method that has minor consequences for the sort of data to which the method has been previously applied (1,2), with 3-4 animals sampled per time point, but which can have more serious consequences when only two animals are sampled per time point. This complication relates to the choice of critical values used in the statistical procedures for hypothesis testing and confidence-interval estimation. That choice is related to the assumption of known variances. In this report we will focus on confidence intervals. In addition to intervals for differences between AUC's, we are also interested in intervals for single AUC's. Such intervals are useful not only to confirm exposure in the individual treatment groups, but also to provide a basis for comparing exposures in the toxicology species with exposures in later human trials.

METHODS

Suppose a study involves J treatment groups in which measurements will be taken at K time points \( t_k \), \( k = 1, \ldots, K \); and that at time point \( t_k \), blood is sampled from \( r_k \) animals in each group. Let \( u_{jk} \) be the measured drug concentration from the \( l \)th animal at time \( t_k \) in the \( j \)th group. Let \( \bar{u}_{jk} \) and \( s_{jk}^2 \) be the sample average and sample variance from the \( r_k \) replicates at time \( t_k \). Let \( \mu_{jk} \) and \( \sigma^2_{jk} \) be the population mean and population variance of which \( \bar{u}_{jk} \) and \( s_{jk}^2 \) are estimates.

Bailer's method is to estimate the mean AUC for the \( j \)th group by applying the trapezoidal rule to the \( \bar{u}_{jk} \)'s:

\[
AUC_j = \sum_{k=1}^{K} \frac{w_{jk} \bar{u}_{jk}}{2}
\]  

where the trapezoidal weights, \( w_{jk} \), are

\[
w_1 = (t_2 - t_1)/2 \\
w_k = (t_{k+1} - t_{k-1})/2 \\
w_K = (t_K - t_{K-1})/2
\]

The variance of \( AUC_j \) is

\[
\sigma^2(AUC_j) = \sum_{k=1}^{K} w_{jk}^2 \sigma^2_{jk}/r_k
\]

an estimator of which is

\[
s^2(AUC_j) = \sum_{k=1}^{K} w_{jk}^2 s^2_{jk}/r_k
\]

1 Departments of Clinical Pharmacology, Drug Safety, Sandoz Research Institute, Sandoz Pharmaceuticals Corporation, 59 Route 10, East Hanover, New Jersey 07936.
2 Department of Drug Metabolism and Pharmacokinetics, Drug Safety, Sandoz Research Institute, Sandoz Pharmaceuticals Corporation, 59 Route 10, East Hanover, New Jersey 07936.
3 To whom correspondence should be addressed.
To construct a confidence interval for the difference of two mean AUC’s, say AUC₁ and AUC₂, Bailer assumed that
\[ s^2(\text{AUC}_1) + s^2(\text{AUC}_2) \]
is in fact the true variance of the difference \( \text{AUC}_1 - \text{AUC}_2 \), rather than just an estimate of
that variance. The resulting confidence interval was of the form
\[ \text{AUC}_1 - \text{AUC}_2 \pm z_{\text{crit}} \sqrt{s^2(\text{AUC}_1) + s^2(\text{AUC}_2)} \] (4)
with a critical value, \( z_{\text{crit}} \), from a standard normal distribution.

Although Bailer did not discuss confidence intervals for a single mean AUC, we will include under the name “Bailer’s method” not only intervals such as (4), but also intervals of the form
\[ \text{AUC} \pm z_{\text{crit}} \sqrt{s^2(\text{AUC})} \] (5)
for a single mean AUC.

Generally, substituting sample variances for population variances is safe when sample sizes are large enough, for then \( t_{\text{crit}} \) approximates \( z_{\text{crit}} \). Here, however, the adequacy of assuming that \( s^2(\text{AUC}) \) is in fact \( \sigma^2(\text{AUC}) \) in (4) and (5) depends not only on the \( r_k \)'s (i.e., on the sample sizes), but also on the \( w_k \)'s and the \( \sigma^2_{jk} \)'s. This is because \( s^2(\text{AUC}) \), as a weighted sum of sample variances, has a complicated distribution that can be approximated as a chi-square with degrees of freedom (df), \( v_j \), given by Satterthwaite’s approximation (4):
\[ v_j = \left( \sum_{k=1}^{K} \frac{w_k^2 \sigma^2_{jk}}{r_k} \right)^2 \left/ \sum_{k=1}^{K} \left( \frac{w_k^4 \sigma_{jk}^4}{r_k^2(r_k - 1)} \right) \right. \] (6a)

It can be demonstrated that
\[ \min(r_k - 1, k = 1, \ldots, K) \leq v_j \leq \sum_{k=1}^{K} (r_k - 1) \] (7)

Moreover, for \( s^2(\text{AUC}) \) the approximating chi-square distribution has a Satterthwaite df of a similar form,
\[ v = \left( \sum_{j=1}^{2} \sum_{k=1}^{K} \frac{w_k^2 \sigma^2_{jk}}{r_k} \right)^2 \left/ \sum_{j=1}^{2} \sum_{k=1}^{K} \left( \frac{w_k^4 \sigma^2_{jk}}{r_k^2(r_k - 1)} \right) \right. \] (6b)

The result can be less than the sum of the two separate df’s.

By the “Bailer-Satterthwaite method” we will mean constructing confidence intervals as in (4) – (5) with \( z_{\text{crit}} \), the critical value assuming known variances, replaced by \( t_{\text{crit}} \), a critical value based on the t-distribution with the Satterthwaite df.

If prior information about the \( \sigma^2_{jk} \)'s is available, the experiment could be designed to improve the chances for a larger df. The form of (6) indicates that more replicates

---

Table I. Nominal 95% Confidence Intervals for Published Data

<table>
<thead>
<tr>
<th>Bailer’s Data&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Bailer-Satterthwaite Method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Low</td>
<td>.036</td>
</tr>
<tr>
<td>Mid</td>
<td>.269</td>
</tr>
<tr>
<td>Hi</td>
<td>.506</td>
</tr>
<tr>
<td>Mid-Mid</td>
<td>.063</td>
</tr>
<tr>
<td>Hi-Mid</td>
<td>.196</td>
</tr>
<tr>
<td>Hi-Low</td>
<td>.009</td>
</tr>
</tbody>
</table>

Yuan’s Data<sup>f</sup>

<table>
<thead>
<tr>
<th>Bailer’s Method</th>
<th>Bailer-Satterthwaite Method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Female</td>
<td>898</td>
</tr>
<tr>
<td>Male</td>
<td>798</td>
</tr>
<tr>
<td>M-F</td>
<td>-152</td>
</tr>
</tbody>
</table>

<sup>a</sup> Pre-phenylmercapturic acid in the livers of mice, \( \mu \text{mole/g} \).
<sup>b</sup> LOW, MID, or HI are dose rates. MID-LOW, HI-MID, and HI-LOW are differences between the respective dose rates.
<sup>c</sup> “Lower” and “Upper” are endpoints of 95% confidence intervals computed for the actual data reported in Bailer’s and Yuan’s publications.
<sup>d</sup> Coverage is the percent coverage from the Monte Carlo simulation described in the text.
<sup>e</sup> Plasma pentachlorophenol concentrations after gavage administration of pentachloroanisole to \( \text{B3C3F1 mice, \( \mu \text{g/mL} \).} \)
<sup>f</sup> M-F is MALE-FEMALE.